(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 024 201 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
02.08.2000 Bulletin 2000/31

(51) Int Cl.[7]: **C12Q 1/68**

(21) Application number: **99400189.9**

(22) Date of filing: **27.01.1999**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(71) Applicants:
- **COMMISSARIAT A L'ENERGIE ATOMIQUE**
 75015 Paris (FR)
- **CENTRE NATIONAL DE**
 LA RECHERCHE SCIENTIFIQUE
 75794 Paris Cédex 16 (FR)

(72) Inventors:
- **Cheval, Lydie**
 91240 Saint Michel sur Orge (FR)
- **Elalouf, Jean-Marc**
 92160 Antony (FR)
- **Virlon, Bérangère**
 92310 Sevres (FR)

(74) Representative: **Vialle-Presles, Marie José et al**
Cabinet Orès,
6, Avenue de Messine
75008 Paris (FR)

Remarks:
The applicant has subsequently filed a sequence listing and declared, that it includes no new matter.

(54) Microassay for serial analysis of gene expression and applications thereof

(57) Method of obtaining a library of tags able to define a specific state of a biological sample, comprising the following successive steps: (1) extracting in a single-step mRNA from a small amount of a biological sample using oligo$(dT)_{25}$ covalently bound to paramagnetic beads, (2) generating a double strand cDNA library, from said mRNA, (3) cleaving the obtained cDNAs using Sau3A I, (4) separating the cleaved cDNAs in two aliquots, (5) ligating the cDNA contained in each of said two aliquots via said Sau3A I restriction site to a linker consisting of one double-strand cDNA molecule having one of the following formulas GATCGTCCC-X_1 or GATCGTCCC-X_2, wherein X_1 and X_2, which comprise 30-37 nucleotides and are different, include a 20-25 bp PCR priming site with a Tm of 55°C-65°C, (6) digesting the products obtained in step (5) with the tagging enzyme BsmF I, (7) blunt-ending said BsmF I tags with a DNA polymerase and mixing the tags ligated with the different linkers, (8) ligating the tags obtained in step (7) to form ditags with a DNA ligase, (9) amplifying the ditags obtained in step (8) with primers comprising 20-25 bp and having a Tm of 55°-65°C, (10) isolating the ditags having between 20 and 28 bp from the amplification products obtained in step (9) by digesting said amplification products with Sau3A I and separating the digested products, (11) ligating the ditags obtained in step (10) to form concatemers, purifiying said concatemers and separating the concatemers having more than 300 bp, (12) cloning and sequencing said concatemers and (13) analysing the different obtained tags.

EP 1 024 201 A1

Description

[0001] Several methods are now available for monitoring gene expression on a genomic scale. These include DNA microarrays (1, 2) and macroarrays (3, 4), expressed sequence tag (EST) determination (5, 6), and serial analysis of gene expression (7). Such methods have been designed, and are still used, for analysing macroamounts of biological material (1-5 µg of poly(A) mRNAs, *i.e.* ~10^7 cells). However, mammalian tissues consist of several different cell types with specific physiological functions and gene expression patterns. Obviously, this makes intricate the interpretation of large scale expression data in higher organisms. It is therefore most desirable to set out methods suitable for the analysis of defined cell populations.

[0002] SAGE has been shown to provide rapid and detailed information on transcript abundance and diversity (7-10). It involves several steps for mRNA purification, cDNA tags generation and isolation, and PCR amplification. We reasoned that increasing the yield of the various extraction procedures, together with slight modifications in the number of PCR cycles could enlarge SAGE potentiality. Here we present a microadaptation of SAGE, referred to as SADE (11) since, in contrast to the original method, it allows to provide quantitative gene expression data on a small number (30,000-50,000) of cells.

[0003] SAGE was first described by Velculescu et al. in 1995 (US Patent 5 695 937 and 7), and rests on 3 principles which have now been all corroborated experimentally: a) short nucleotide sequence tags (10 bp) are long enough to be specific of a transcript, especially if they are isolated from a defined portion of each transcript; b) concatenation of several tags within a single DNA molecule greatly increases the throughput of data acquisition; c) the quantitative recovery of transcript-specific tags allows to establish representative gene expression profiles.

[0004] However, said method was designed to study macroamounts of biological materials (5 µg of poly(A) RNAs, i.e. about 10^7 cells). Since mammalian tissues consist of several different cell types with specific physiological functions and gene expression patterns, it is most desirable to scale down the SAGE approach for studying well delineated tissue fragments or isolated cell populations.

[0005] The inventors have now found a new method able to handle microamounts of samples.

[0006] The subject of the present invention is a method of obtaining a library of tags able to define a specific state of a biological sample, such as a tissue or a cell culture, characterised in that it comprises the following successive steps:

(1) extracting in a single-step mRNA from a small amount of a biological sample using oligo$(dT)_{25}$ covalently bound to paramagnetic beads,
(2) generating a double-strand cDNA library, from said mRNA according to the following steps:

- synthesising the 1st strand of said cDNA by reverse transcription of said mRNA template into a 1st complementary single-strand cDNA, using a reverse transcriptase lacking Rnase H activity,
- synthesising the 2nd strand of said cDNA by nick translation of the mRNA, in the mRNA-cDNA hybrid form by an *E. coli* DNA polymerase,

(3) cleaving the obtained cDNAs using the restriction endonuclease Sau3A I as anchoring enzyme,
(4) separating the cleaved cDNAs in two aliquots,
(5) ligating the cDNA contained in each of said two aliquots via said Sau3A I restriction site to a linker consisting of one double-strand cDNA molecule having one of the following formulas:

$$GATCGTCCC\text{-}X_1 \text{ or } GATCGTCCC\text{-}X_2,$$

wherein X_1 and X_2, which comprise 30-37 nucleotides and are different, include a 20-25 bp PCR priming site with a Tm of 55°C-65°C, and
wherein GATCGTCCC correspond to a Sau3A I restriction site joined to a BsmF I restriction site,

(6) digesting the products obtained in step (5) with the tagging enzyme BsmF I and releasing linkers with anchored short piece of cDNA corresponding to a transcript-specific tag, said digestion generating BsmF I tags specific of the initial mRNA,
(7) blunt-ending said BsmF I tags with a DNA polymerase, preferably T7 DNA polymerase or Vent polymerase and mixing the tags ligated with the different linkers,
(8) ligating the tags obtained in step (7) to form ditags with a DNA ligase,
(9) amplifying the ditags obtained in step (8) with primers comprising 20-25 bp and having a Tm of 55°-65°C,
(10) isolating the ditags having between 20 and 28 bp from the amplification products obtained in step (9) by digesting said amplification products with the anchoring enzyme Sau3A I and separating the digested products on

an appropriate gel electrophoresis,
(11) ligating the ditags obtained in step (10) to form concatemers, purifiying said concatemers and separating the concatemers having more than 300 bp,
(12) cloning and sequencing said concatemers and
(13) analysing the different obtained tags.

[0007] According to an advantageous embodiment of said method, in step (2), said synthesis of the 1st strand of said cDNA is performed with Moloney Murine Leukaemia Virus reverse transcriptase (M-MLV RT), and oligo$(dT)_{25}$ as primers.

[0008] According to another advantageous embodiment of said method, the linkers of step (5) are preferably hybrid DNA molecules formed from linkers 1A and 1B or from linkers 2A and 2B, having the following formulas:

linker 1A: 5'-TTTTGCCAGGTCACTCAAGTCGGTCATTCATGTCAGCACAGGGAC-3'

linker 1B: 5'-GATCGTCCCTGTGCTGACATGAATGACCGACTTGAGTGACCTGGCA-3'

or

linker 2A: 5'-TTTTTGCTCAGGCTCAAGGCTCGTCTAATCACAGTCGGAAGGGAC-3'

linker 2B: 5'-GATCGTCCCTTCCGACTGTGATTAGACGAGCCTTGAGCCTGAGCAA-3'.

[0009] According to another advantageous embodiment of said method, the amount of each linker in step (5) is at most of 8-10 pmol and preferably comprised between 0.5 pmol and 8 pmol for initial amounts of respectively 10-40 ng of mRNAs and 5 µg of mRNAs.

[0010] According to yet another advantageous embodiment of said method, the primers of step (9) have preferably the following formulas:

5'-GCCAGGTCACTCAAGTCGGTCATT-3'

5'-TGCTCAGGCTCAAGGCTCGTCTA-3'.

[0011] According to yet another advantageous embodiment of said method, the biological sample of step (1) preferably, comprises ≤ 5.10^6 cells, corresponding to at most 50 µg of total RNA or 1 µg of poly(A) RNA.

[0012] According to the invention, biological sample means for instance : tissue, cells (native or cultured cells), which are lysed for extracting mRNA.

[0013] According to another advantageous embodiment of said method, said tissue sample is from kidney, more specifically from nephron segments corresponding to about 15,000 to 45,000 cells, corresponding to 0.15-0.45 µg of total RNA.

[0014] The subject of the present invention is also the use of a library of tags obtained according to the method as defined above, for assessing the state of a biological sample, such as a tissue or a cell culture.

[0015] The subject of the present is also the use of the tags obtained according to the method as defined here above as probes.

[0016] The subject of the present invention is also a method of determination of a gene expression profile, characterised in that it comprises :

- performing steps (1) to (13) according to claim 1 and
- translating cDNA tag abundance in gene expression profile.

[0017] According to a preferred embodiment of said method the gene expression profile obtained in mouse outer medullary collecting duct (OMCD) and in mouse medullary thick ascending limb (MTAL) is as specified in Table I below:

OMCD	MTAL	Tag	GenBank match
99	2	GTGGCAGTGG	EST (AA097074) similar to rat AQP-2 (D13906)
34	1	TTATAATTTG	ESTs
27	0	TGGCAGTGGG	No match
19	5	TGACTCCCTC	B2 repetitive sequence
13	0	AAGTTTAAAT	Thymosin beta-4 (X16053)
13	1	AGCAAGCAGG	β-actin (X03672)
13	4	CAAAAAGCTA	ESTs, similar to rat ribosomal protein L11 (X62146)
11	1	ACATTCCTTA	ESTs
11	14	ACCGACCGCA	Integral membrane protein 2B1 (U76253)
11	0	CAGAAGAAGT	Endogenous murine leukemia virus (M17326)
10	5	AAATAAAGTT	Lactate deshydrogenase 2, B chain (X51905)
10	0	AGAAGCAGTG	EST 750555 (AA472938)
10	6	TGATGCCCTC	B2 repetitive sequence
9	4	AGGCTACTAC	Ribosomal protein L27a (X05021)
9	11	GCTCATTGGA	ESTs
9	6	GCTTTCAGCA	ESTs, similar to human extracellular proteinase inhibitor hom (X63187)
9	14	GTGACTGGGT	CytC oxidase subunit IV (X54691)
9	0	TGACCAAGGC	11 β-hydroxysteroid dehydrogenase type 2 (X90647)

[0018] The invention also relates to a kit useful for detection of gene expression profile, characterised in that the presence of a cDNA tag, obtained from the mRNA extracted from a biological sample, is indicative of expression of a gene having said tag sequence at an appropriate position, i.e. immediately adjacent to the most 3' Sau3A I site in said cDNA, obtained form said mRNA, the kit comprising further to usual buffers for cDNA synthesis, restriction enzyme digestion, ligation and amplification,

- containers containing a linker consisting of one double-strand cDNA molecule having one of the following formulas:

$$GATCGTCCC\text{-}X_1 \text{ or } GATCGTCCC\text{-}X_2,$$

wherein X_1 and X_2, which comprise 30-37 nucleotides and are different, include a 20-25 bp PCR priming site with a Tm of 55°C-65°C, and
wherein GATCGTCCC correspond to a Sau3A I restriction site joined to a BsmF I restriction site, and

- containers containing primers comprising 20-25 bp and having a Tm of 55°-65°C

[0019] According to an advantageous embodiment of said kit, it preferably contains

- containers containing hybrid DNA molecules formed from linkers 1A and 1B or from linkers 2A and 2B, having the following formulas:

linker 1A: 5'-TTTTGCCAGGTCACTCAAGTCGGTCATTCATGTCAGCACAGGGAC-3'

linker 1B: 5'-GATCGTCCCTGTGCTGACATGAATGACCGACTTGAGTGACCTGGCA-3'

or

linker 2A: 5'-TTTTTGCTCAGGCTCAAGGCTCGTCTAATCACAGTCGGAAGGGAC-3'

linker 2B: 5'-GATCGTCCCTTCCGACTGTGATTAGACGAGCCTTGAGCCTGAGCAA-3',

and

- containers containing the following primers:

5'-GCCAGGTCACTCAAGTCGGTCATT-3'

5'-TGCTCAGGCTCAAGGCTCGTCTA-3'.

[0020] As compared to SAGE, the instant SADE method includes the folowing features: 1) single-step mRNA purification from tissue lysate; 2) use of a reverse transcriptase lacking Rnase H activity; 3) use of a different anchoring enzyme; 4) modification of procedures for blunt-ending cDNA tags; 5) design of new linkers and PCR primers.

[0021] Figure 1, modified from the original studies of Velculescu et al., summarises the different steps of the SADE method, which is a microadaptation of SAGE. Briefly, as already specified here above, mRNAs are extracted using oligo$(dT)_{25}$ covalently bound to paramagnetic beads. Double strand cDNA is synthesised from mRNA using oligo$(dT)_{25}$ as primer for the 1 st strand synthesis. The cDNA is then cleaved using a restriction endonuclease (*anchoring enzyme:* SAU3A I) with a 4-bp recognition site. Since such an enzyme cleaves DNA molecules every 256 bp (4^4) on average, virtually all cDNAs are predicted to be cleaved at least once. The 3' end of each cDNA is isolated using the property of the paramagnetic beads and divided in half. Each of the two aliquots is ligated via the anchoring enzyme restriction site to one of the two linkers containing a type IIS recognition site (*tagging enzyme*: BsmF I) and a priming site for PCR amplification. Type IIS restriction endonucleases display recognition and cleavage sites separated by a defined length (14 bp for BsmF I), irrespective of the intercalated sequence. Digestion with the type IIS restriction enzyme thus releases linkers with an anchored short piece of cDNA, corresponding to a transcript-specific tag. After blunt ending of tags, the two aliquots are linked together and amplified by PCR. Since all targets are of the same length (110 bp) and are amplified with the same primers, potential distortions introduced by PCR are greatly reduced. Furthermore, these distortions can be evaluated, and the data corrected accordingly (7, 8). Ditags present in the PCR products are recovered through digestion with the anchoring enzyme and gel purification, then concatenated and cloned.

[0022] In the SAGE method, mRNAs are isolated using conventional methods, then hybridized to biotinylated oligo (dT) for cDNA synthesis. After cleavage with the anchoring enzyme Nla III, the biotionylated cDNA fraction (3' end) is purified by binding to streptavidin beads.

[0023] In the SADE method, mRNAs are directly isolated from the tissue lysate through hybridization to oligo(dT) covalently bound to magnetic beads. Then, all steps of the experiment (until step 3 of protocol 5, as described here after) are performed on magnetic beads. This procedure saves time for the initial part of the experiment and, more importantly, provides better recovery. Quantitative analysis of the cDNA amounts available for library construction revealed dramatic differences between SAGE and SADE. With the SAGE method, starting from 500 mg tissue, 1.7 µg of cDNA are obtained, and only 4 ng were able to bind to streptavidin beads after Sau3A I digestion. With the SADE method, starting from 250 mg of tissue, 3.2 µg of cDNA were synthesised on beads, and 0.5 µg remained bound after Sau3A I cleavage. The increased yield of SADE (X250) explains the success in constructing libraries from as few as 30,000 cells. Using different sources of oligo(dT) leaded to poor cDNA recoveries. This may be explained by the fact that the binding capacity of streptavidin beads can be altered by several parameters, such as the presence of phenol, the length and composition of the biotinylated DNA fragments, and the length of the spacer between the oligo(dT) and the biotin molecule.

[0024] Another important difference between SAGE and SADE concerns the selected anchoring enzyme. Although any restriction enzyme with a 4-bp restriction site could serve as anchoring enzyme, Sau3A I was preferred to Nla III (7-10) or other enzymes in our studies. Several cDNA libraries used for large scale sequencing are constructed by vector priming, followed by cDNA cleavage with Mbo I (an isoschizomer of Sau3A I which does not cut the vector (methylated) DNA), and circularisation (6). SADE tags therefore correspond to the cDNA 5' ends of these libraries, which enables to use more efficiently EST data bases to analyse the data.

[0025] In addition to the preceding arrangements, the invention further comprises other arrangements, which will emerge from the description which follows, which refers to examples for carrying out the process which is the subject of the present invention as well as to the accompanying drawings, in which:

- **Figure 1:** Outline of procedures for constructing SADE libraries. Poly(A) RNAs are isolated from tissue lysate using oligo$(dT)_{25}$ covalently linked to paramagnetic beads, and cDNA is synthesised under solid-phase condition. Bold face characters correspond to biologically relevant sequences, whereas light characters represent linker-

derived sequences. The anchoring enzyme (AE) is Sau3A I, whereas the tagging enzyme (TE) is BsmF I. See text for details.

- **Figure 2:** Gel analysis of cDNAs synthesised from different amounts of tissue. Poly(A) RNAs were isolated from the indicated amounts of mouse kidney, and cDNAs were synthesised and Sau3A I-digested on paramagnetic beads (see protocols 1-2). cDNAs released from beads were recovered, and half of the material obtained from each reaction was analysed on a 1% agarose gel stained with ethidium bromide. Position of molecular weight markers are indicated in bp: left, λ Bste II-digest; right, pBR Msp I-digest.
- **Figure 3:** PCR amplification of ditags. Poly(A) RNAs were isolated from 50 or 150 mm of microdissected nephron segments (corresponding to about 15,000 and 45,000 cells, respectively). The corresponding ditags were amplified by PCR using the indicated number of cycles and analysed on a 3% agarose gel stained with ethidium bromide. The expected product (linkers + 1 ditag) is 110-bp long. Molecular weight marker (M) is 10-bp DNA ladder (Life Technologies).
- **Figure 4:** Gel analysis of concatemers. Ditags were concatenated by ligation (2 h at 16°C), then electrophoresed through an 8% polyacrylamide gel. The gel was post-stained using SYBR Green I, and visualised by UV illumination at 305 nm. Migration of the molecular weight marker (λ BstE II-digest) is indicated on the right.
- **Figure 5:** Comparison of gene expression levels in two nephron portions of the mouse kidney. SADE libraries were constructed from ~50,000 cells isolated by microdissection from medullary collecting ducts or medullary thick ascending limbs, and 5,000 tags were sequenced in each case. The data show the 18 most abundant collecting duct tags originating from nuclear transcripts (mitochondrial tags were excluded from the analysis), and their corresponding abundance in the thick ascending limb library.

[0026] It should be understood, however, that these examples are given solely by way of illustration of the subject of the invention and do not in any way constitute a limitation thereto.

Example:

1. Tissue sampling and mRNA isolation

1.1 Tissue sampling and lysis

[0027] The initial steps of library construction require the usual precautions recommended for experiments carried out with RNAs (12). In addition, since library construction involves large scale PCR (*Protocol 6*), care must be taken to avoid contamination from previous libraries. Working under PCR grade conditions is especially important when low amounts of tissue or cells are used.

[0028] Starting from whole tissues (*i.e.* kidney, liver, brain, ...), the following procedures may be routinely used. After animal anaesthesia or decapitation, the tissue is removed as quickly as possible, rapidly rinsed in ice-cold phosphate-buffered saline, sliced in ~50 mg-pieces, and frozen in liquid nitrogen. The frozen sample is then ground to a fine powder under liquid nitrogen using a mortar and a pestle, transferred into lysis binding buffer (*protocol 1*), and homogenised with a Dounce tissue disrupter. To avoid loss of material, small samples (≤20 mg) can be transferred without previous freezing in the lysis binding buffer, and homogenised in a 1 ml Dounce. The respective amounts of tissue and lysis binding buffer needed for a variety of conditions are indicated in Table II.

Table II.

Small and large scale mRNA isolation and cDNA synthesis				
Tissue/ cells	Lysis binding buffer (ml)	Oligo(dT) beads (µl)	Reaction volume (µl)	
			1st strand	2nd strand
250 mg/ $3x10^7$	5.50	600	50	400
30 mg/$3x10^6$-$6x10^6$	0.70	100	50	400
4 mg/10^5-10^6	0.10	30	25	200
0.5 mg/$3x10^4$-10^5	0.05-0.10	20	25	200

[0029] Starting from isolated or cultured cells, the procedure is much more rapid. The cell suspension, maintained in appropriate culture or survival medium, just needs to be centrifuged at 600-1,200g for 5 min. After supernatant removal, the lysis binding buffer is added onto the cell pellet, and the sample is homogenised by vortexing. This procedure has been successfully applied to $3X10^4$-$3X10^7$ cells (Table II).

1.2. mRNA isolation.

[0030] Protocols 1-7 describe the generation of a SADE library from 0.5 mg of tissue. The amount of cDNA recovered corresponds to an experiment carried out on the mouse kidney. Slightly different amounts are expected to be obtained from other tissues, according to their mRNA content. The procedures described herein have been repeatedly used without modifications with $3x10^4$-10^5 isolated cells. Since some applications can be performed on large amounts of tissue or cells, protocol adaptations and anticipated results for these kinds of experiments are also provided.

[0031] In the initial experiments, RNAs were extracted using standard methods (13), and poly(A) RNAs were isolated on oligo(dT) columns. Besides being time consuming, this procedure provides low and variable mRNA amounts, and cannot be easily scaled down. The alternative procedure described here (use of oligo$(dT)_{25}$ covalently linked to paramagnetic beads) is a single tube assay for mRNA isolation from tissue lysate. In our hands, it yields 4-times higher mRNA amounts than standard methods. Kits and helpful instructions for mRNA isolation with oligo(dT) beads can be obtained from Dynal. Handling of these beads is relatively simple, but care must be taken to avoid centrifugation, drying or freezing, since all three processes are expected to lower their binding capacity. On the other hand, beads can be resuspended by gentle vortexing or pipetting without extreme precautions.

Protocol 1. mRNA purification

[0032] *Equipment and reagents*

- Appropriate tissue or cells.
- Dynabead mRNA direct kit (Dynal, ref. 610-11) containing Dynabeads oligo $(dT)_{25}$, lysis binding buffer, and washing buffers.
- 5X reverse transcription (RT) buffer (250 mM Tris-HCl (pH8.3), 375 mM KCl, 15 mM $MgCl_2$), provided with cDNA synthesis kit (see *protocol 2*).
- Magnetic Particle Concentrator (MPC) for 1.5 ml tubes (Dynal, ref. 12004).
- Glycogen for molecular biology (Boehringer Mannheim, ref. 901393).

[0033] *Method*

1. Lyse the tissue sample in 100 µl lysis binding buffer supplemented with 10 µg glycogen.
2. Add 20 µl of Dynabeads in a 1.5 ml tube and condition them according to manufacturer's instructions.
3. Using the MPC, remove the supernatant from the Dynabeads and add the tissue lysate (100 µl). Mix by vortexing and anneal mRNAs to the beads by incubating 10 min at room temperature.
4. Place the tube 2 to 5 min in the MPC and remove the supernatant. The mRNAs are fixed on the beads.
5. Using the MPC, perform the following washes (all buffers contain 20 µg/ ml glycogen): twice with 200 µl washing buffer containing lithium dodecyl sulfate (LiDS), 3-times with 200 µl washing buffer, and twice with 200 µl ice-cold 1X RT buffer. Resuspend the beads by pipetting, transfer the suspension in a fresh 1.5 ml tube, wash once with 200 µl ice-cold 1X RT buffer and immediately proceed to *protocol 2*. mRNAs on the beads are now ready for 1 st strand cDNA synthesis.

1.3 mRNA integrity and purity

[0034] Before generating a cDNA library, it is generally advised to check for mRNA integrity by Northern blot analysis. However, this control experiments consumes part of the material, takes several days, and often leads to ambiguous results (a variety of reasons can cause poor Northern hybridisation signals). In addition, it is no longer possible when using small amounts of tissue or cells. RNA degradation has only to be expected in the three following conditions: 1) cell survival is not maintained before lysis or freezing; 2) cell thawing outside of lysis buffer, and 3) use of poor quality reagents. Since Rnase-free reagents are now available from a variety of company, it is much more rapid and effective to check for survival (*i.e.* select the appropriate culture medium) and freezing conditions than to perform tricky tests on RNA aliquots.

[0035] The purity of mRNAs isolated with oligo(dT) beads is better than that obtained with conventional methods. When we generated SAGE libraries using mRNAs extracted with guanidinium thiocyanate and oligo(dT) columns, nuclear encoded rRNAs amounted to 1% of the sequenced tags. Using the alternative mRNA extraction procedure, rRNAs tags are no longer present in the library.

2. 1st and 2nd strand synthesis

2.1. 1st strand cDNA synthesis

[0036] The first step in the synthesis of cDNA is copying the mRNA template into complementary single-strand cDNA. In *protocol 2,* 1 st strand cDNA is synthesised using Moloney Murine Leukaemia Virus RT (M-MLV RT). With this enzyme, we have been able to generate SADE libraries from either large or minute amounts of cells (Table 1). In our last series of experiments, we have however used Superscript II M-MLV RT, provided with the Superscript cDNA synthesis kit (Life Technologies, ref. 18090-019). In this case, the amount of cDNA formed (see 2.3) was increased ~4-fold. Although this better yield likely results from both the synthesis of longer cDNAs (which is not essential for the current application) and of a higher number of cDNA molecules, we strongly recommend to use Superscript II M-MLV RT for very small samples (≤50,000 cells). The protocol will be similar to the one described here, except for reaction volumes (20 µl for 1st strand synthesis, and 150 µl for 2nd strand synthesis).

[0037] mRNAs are generally heated 5 min at 65°C before reverse transcription to break up secondary structures. Since such a high temperature will also denature the mRNA-oligo$(dT)_{25}$ hybrid, we only heat the sample at 42°C before initiation of 1st strand synthesis.

2.2. 2nd strand synthesis

[0038] Many procedures have been developed for 2nd strand cDNA synthesis. The method used here is a modification of the Grubler and Hoffman procedure. Briefly, the mRNA (in the mRNA-cDNA hybrid) is nicked by *E. coli* RNAse H. *E. coli* DNA polymerase initiates the second strand synthesis by nick translation. *E. coli* DNA ligase seals any breaks left in the second strand cDNA. The procedure is described in *protocol 2.* This step is usually very efficient (approximately 100%) so that a 2 h-incubation period is sufficient when starting from macroamounts of material (>100 mg of tissue or 10^7 cells).

Protocol 2. cDNA synthesis and cleavage

[0039] *Equipment and reagents*

- cDNA synthesis kit (Life Technologies, ref. 18267-013) contains all buffers and enzymes necessary for first and second strand cDNA synthesis.
- α[^{32}P]dCTP 6000Ci/mmol (Amersham, ref. AA0075).
- TEN (10 mM Tris-HCl (pH8.0), 1 mM EDTA, 1 M NaCl).
- Restriction endonuclease Sau3A I 4 U/µl (New England Biolabs, ref. 169L), provided with 10X reaction buffer and purified 100X bovine serumalbumin (BSA, 10 mg/ ml).
- Magnetic Particle Concentrator MPC (Dynal).
- Geiger counter.
- Automated thermal cycler or water-baths equilibrated at 42°C, 37°C, and 16°C.

[0040] *Method*

1. Resuspend the beads in 12.5 µl of 1X first strand (*i.e.* RT) reaction buffer.
2. Incubate 2 min at 42°C.
3. Place the tube at 37°C for 2 min. Add 12.5 µl of the following mix : 5 µl DEPC-treated water, 2.5 µl 5X first strand buffer, 1.25 µl dNTP 10 mM, 2.5 µl DTT 100 mM, 1.25 µl MMLV reverse transcriptase.
3. Incubate 1 h at 37°C and chill on ice.
4. On ice, prepare the following mix : 169.7 µl DEPC-treated water, 4.5 µl dNTP 10 mM, 24 µl 2nd strand buffer, 2 µl α[^{32}P]dCTP, 6 µl *E. coli* DNA polymerase I, 1.05 µl *E. coli* RNAse H, 0.75 µl *E. coli* DNA ligase, 2 µl glycogen 5 µg/ µl.
5. Add 175 µl to the first strand tube and incubate overnight at 16°C. Keep the remaining mix for subsequent measurement of its radioactivity and calculation of dCTP specific activity.
6. Wash beads to remove non incorporated α[^{32}P]dCTP: 4-times with 200 µl TEN + BSA[a], and 3-times with 200 µl ice-cold 1X mix Sau3A I + BSA[a]. Check with Geiger counter that the last eluate is not radioactive, whereas the material bound on the beads is highly radioactive.
7. Add on the beads the following mix : 88 µl H_2O, 10 µl 10X mix Sau3A I, 1 µl 100X BSA, 1 µl Sau3A I. Incubate 2h at 37°C. Vortex intermittently.

[a]Final concentration of BSA: 0.1 mg/ ml.

8. Chill 5 min on ice.
9. Using the MPC, remove the supernatant, which contains the 5' end of the cDNA. Wash once with 200 µl of 1X mix Sau3A I + BSA[a]. Remove this second supernatant, pool it with the first one, and store the resulting solution (300 µl) in order to measure the yield of second strand synthesis (see text). Before going to *step 10,* check with Geiger counter that both the eluate and beads are radioactive.
10. Resuspend the beads in 200 µl TEN supplemented with BSA[a].

2.3. Yield of 2nd strand synthesis

[0041] A method to calculate the yield for first and second strand cDNA synthesis is given in the cDNA synthesis kit instruction manual. We do not measure the yield of 1st strand cDNA synthesis since, as discussed above (*1.3*), this implies to set away part of the preparation.

[0042] The amount of double strand (ds) cDNA formed is calculated by measuring radioactivity incorporation in the 5' end of the cDNA, which is released in the supernatant after Sau3A I digestion (see *Protocol 2).* The 300 µl-supernatant is extracted with PCI and the ds cDNA is ethanol precipitated in the presence of glycogen (50 µg/ ml) and 2.5 M ammonium acetate. The pellet is resuspended in 8 µl of TE. Half of the material is used for liquid scintillation counting, and the remaining is loaded on a 1.0 or 1.5% agarose gel. For experiments carried out on 250, 30, 4, and 0.5 mg of mouse kidney, we obtained the following amounts (µg) of ds cDNA: 2.8, 0.3, 0.05, and 0.01. The higher amount corresponds to the incorporation of 1.3% of the input radioactivity. In these experiments, three of the four cDNA samples could be detected by ethidium bromide staining after gel electrophoresis (Fig. 2). Their size ranged between <0.2 and ~3 kbp (the small size of most cDNA fragments is due to Sau3A I digestion). When cDNA amounts are below the detection threshold of the ethidium bromide staining method, autoradiographic analysis can be performed. In this case, the gel is fixed in 10% acetic acid, vacuum dried and exposed overnight at-80°C with one intensifying screen for autoradiography.

3. Linkers design, preparation, and ligation

3.1. Linkers design

[0043] A variety of linkers can be used at this point. Linkers must contain three important sequences : a) the appropriate anchoring enzyme overhang; b) a recognition site for a type IIs restriction enzyme (tagging enzyme); c) a priming site for PCR amplification. High quality linkers are crucial for successful library generation.

[0044] Table III provides the sequence of linkers and PCR primers used in our experiments. All four linkers must be obtained gel-purified. Linkers 1B and 2B display two modifications: a) 5' end phosphorylation, and b) C7 amino modification on the 3' end. Linkers phosphorylation can be performed either enzymatically with T4 polynucleotide kinase, or chemically at the time of oligonucleotide synthesis. In both cases, phosphorylation efficiency must be tested *(Protocol 3).* We use chemically phosphorylated linkers. Linkers modification on the 3' end serves to increase the efficiency of ditag formation (*protocol 5, step 8-11*). Indeed, the modified 3' end cannot be blunt-ended and will not ligate to cDNA tags or linkers.

Table III. Sequence of linkers and PCR primers

Oligonucleotide	Sequence
Linker 1A	5′-TTTTGCCAGGTCACTCAAGTCGGTCATTCATGTCAGCACAGGGAC-3′
Linker 1B[a]	5′-GATCGTCCCTGTGCTGACATGAATGACCGACTTGAGTGACCTGGCA-3′
Linker 2A	5′-TTTTTGCTCAGGCTCAAGGCTCGTCTAATCACAGTCGGAAGGGAC-3′
Linker 2B[a]	5′-GATCGTCCCTTCCGACTGTGATTAGACGAGCCTTGAGCCTGAGCAA-3′
Primer 1	5′-GCCAGGTCACTCAAGTCGGTCATT-3′
Primer 2	5′-TGCTCAGGCTCAAGGCTCGTCTA-3′

[a] Linkers 1B and 2B include two modifications (5`-phosphorylation and 3`-C7 amino modification).

[0045] With regard to the PCR priming site, it was designed with the help of Oligo™ software (Medprobe, Norway) in order to obtain PCR primers with high Tm (60°C), and avoid self-priming or sense/ antisense dimer formation. Two different priming sites must be designed in " left " and " right " linkers, otherwise the target will undergo panhandle formation, and thus escape PCR amplification.

Protocol 3. Preparing and testing linkers

[0046] *Equipment and reagents*

- Linkers 1A, 1B, 2A, and 2B at 20 pmol/ µl.
- Primers 1 and 2 at 20 pmol/ µl.
- T_4 DNA ligase 1 U/ µl (Life Technologies, ref. 15224-017) and 5X reaction buffer.
- 10 mM ATP.
- PCR reagents: *Taq* DNA polymerase 5 U/ µl (Eurobio), 10X PCR buffer (200 mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$, 500 mM KCl, 1 mg/ ml gelatin), 1.25 mM dNTP, 100 mM $MgCl_2$, and 100 mM DTT.
- Restriction endonucleases Sau3A I (4 U/ µl) and BsmF I (2 U/ µl) (New England Biolabs, refs. 169L and 572L), provided with 10X reaction buffer and 100X BSA.
- 10-bp DNA ladder (Life Technologies, ref. 10821-015)
- Automated thermal cycler and water baths equilibrated at 14°C, 37°C, and 65°C.
- Tris-HCl buffered (pH 7.9) phenol-chloroform-isoamyl alcool (PCI).
- 10 M ammonium acetate.
- TE (10 mM Tris-HCl (pH8.0), EDTA 1 mM).

[0047] *Method*

1. Mix 25 µl of linker 1A and 25 µl of linker 1B in a 0.5 ml PCR tube (final concentration: 10 pmol/ µl). Proceed similarly for linkers 2A and 2B.
2. Transfer PCR tubes in the thermal cycler. Heat at 95°C for 2 min, then let cool at room temperature for 20 min on the bench. Store at -20°C.
3. Test self-ligation of each hybrid , as well as ligation of hybrid (1A/1B) with hybrid (2A/2B). Set-up 3 ligation

reactions by mixing 1 µl of hybrid (1A/ 1 B) (tube 1), 1 µl of hybrid (2A/ 2B) (tube 2), 0.5 µl of hybrid (1A/1B) and 0.5 µl of hybrid (2A/ 2B) (tube 3) with 2 µl 10 mM ATP, 4 µl 5X ligase mix, 12 µl H2O, and 1 µl T4 DNA ligase.

4. Incubate 2 h or overnight at 14°C. Analyse 10-µl aliquots on a 3% agarose gel using 10-bp DNA ladder as marker. Most of the material (≥80%) consists of a 94-bp DNA fragment.

5. Proceed to PCR using 10^5 targets from tube 3 reaction (dilute using TE buffer supplemented with 0.1 mg/ ml BSA). Mix 1 µl of diluted ligation product, 5 µl 10X PCR buffer, 1 µl 100 mM $MgCl_2$, 8 µl 1.25 mM dNTP, 2.5 µl primer 1, 2.5 µl primer 2, and 30 µl water. Prepare 4 such reactions and a control tube without linker, transfer in the thermal cycler and heat at 80°C for 2 min.

6. Add in each tube 50 µl of *Taq* polymerase amplification mix (5 µl 10X PCR buffer, 4 µl 100 mM DTT, 0.5 µl *Taq* polymerase, 40.5 µl water), and 60 µl of mineral oil if necessary for your thermal reactor.

7. Perform 29 PCR cycles (95°C, 30 s; 58°C, 30 s; 70°C, 45 s), followed by an additional cycle with a 5-min elongation time.

8. Analyse 10-µl aliquots on a 3% agarose gel. A 90-bp amplification fragment is clearly visible.

9. Pool all 4 PCR samples, extract with equal volume of PCI. Transfer the aqueous (upper) phase in a fresh tube, then add 100 µl 10 M ammonium acetate and 500 µl isopropanol. Round the tubes several times for mixing, centrifuge (15,000g) at 4°C for 20 min, wash twice with 400 µl 75% ethanol, vacuum dry, and resuspend the pellet in 12 µl TE buffer.

10. Set-up two 50-µl digestion reactions using 5 µl of DNA and 4 U of Sau3A I or BsmF I. Incubate 1 h at 37°C or 65°C, as appropriate.

11. Analyse 10-µl aliquots on a 3% agarose gel. Run in parallel 1 µl of uncut PCR product. Sau3A I and BsmF I digestion must be completed to ≥ 80%.

3.2. Linkers preparation

[0048] It is essential to check that ds linkers can be ligated, PCR amplified, and digested with the anchoring and tagging enzyme. Success with *protocol 3* experiments is a prerequisite before attempting to prepare a library. The PCR conditions described here have been optimised for Hybaid thermal reactors (TR1 and Touch Down) working under control or simulated tube conditions. Different conditions may be used with other machines. Note that since the target is quite small (90 bp), elongation is performed at a relatively low temperature.

3.3. Ligating linkers to cDNA

[0049] The concentration of ds linkers should be adapted to the amount of cDNA used to prepare the library. In the original protocol of Velculescu *et al.*, 2 µg (74 pmol) of ds linkers are used. Considering that starting from 5 µg of mRNAs, 2 µg of cDNAs with 1-2 kb average size are obtained, the amount of cDNA available for ligation is in the range of 1.5-3 pmol. Since a large excess of linkers decreases the PCR signal to noise ratio, we perform ligation with 8 pmol ds linkers for libraries generated from 250 mg of tissue (~5 µg mRNAs). Starting from $5X10^4$-10^5 cells (10-40 ng mRNAs), 0.5 pmol of ds linkers are used. A lower amount of linkers may allow efficient ligation, but we have no experience for it.

Protocol 4. Ligating ds linkers to cDNA

[0050] *Equipment and Reagents*

- Hybrid (1A/1B) and (2A/2B) at 0.5 pmol/ µl, obtained from *protocol 3*, steps 1-2.
- TEN, TE, and LoTE (3 mM Tris-HCl (pH7.5), 0.2 mM EDTA), stored at 4°C.
- 10X NEB IV reaction buffer and 100X BSA (New England Biolabs).
- T_4 DNA ligase 5 U/ µl (Life Technologies, ref. 15224-041) and 5X ligation mix; 10 mM ATP.
- MPC (Dynal).
- Water-baths equilibrated at 45°C and 16°C.
- Geiger counter

[0051] *Method*

1. Once the experiments described in *protocol 2* have been carried out, perform 2 additional washes of the beads before ligating ds linkers to the cDNA. Using the MPC, wash the beads with 200 µl of TEN + BSA.[a] Resuspend the beads in 200 µl of the same buffer (take care to recover the beads completely: mix by repeated pipetting and

[a] Final concentration of BSA: 0.1 mg/ ml.

scrape the tube wall with the pipette tip), then separate into two 100 µl-aliquots: one will be ligated to hybrid (1A/1B), the other will be ligated to hybrid (2A/2B).

2. Add 10 µl of fresh Dynabeads in two 1.5 ml tubes. These tubes will be now treated as the two others and will be used as negative control.

3. Wash twice the 4 tubes with 200 µl of ice-cold TE buffer + BSA[a].

4. Immediately after the last rinsing, add to each tube 34 µl of the appropriate mix containing 8 µl of 5X ligase buffer and 0.5 pmol of hybrid (1A/1B) or hybrid (2A/2B). Heat 5 min at 45°C then chill on ice.

5. Add in each tube 4 µl of 10 mM ATP, and 2 µl of T_4 DNA ligase (final volume: 40 µl). Incubate overnight at 14°C.

6. Wash beads thoroughly (free linkers will poison the PCR amplification) as follows: 4-times with 200 µl of TEN + BSA[a] and 3-times with 200 µl of 1X NEBIV + BSA[a]. After the first rinsing with NEB IV, take care to resuspend completely the beads (see above) and transfer them to fresh tubes. After the last rinsing, check that radioactivity is still present on the beads, but absent from the supernatant.

7. Proceed to *protocol 5* or store at 4°C.

[0052] After ligation (step 5 in *protocol 4),* it is very important to wash the beads extensively in order to remove free ds linkers. In fact, if ds linkers not ligated to cDNA fragments are not thoroughly eliminated from each sample, the library will contain large amounts (up to 25%) of linkers sequences. This will make data acquisition poorly efficient.

4. Ditags formation

4.1. Release of cDNA tags

[0053] Digestion with the tagging enzyme (BsmF I) will release only small DNA fragments from oligo(dT) beads. Consequently, much of the radioactivity remains bound to the beads at this stage. In order to check that extensive rinsing did not cause great loss of material, we usually measure beads radioactivity by Cerenkov counting (the data must be corrected for the efficiency of Cerenkov counting (~50 % of liquid scintillation counting efficiency) after BsmF I digestion. For experiments previously described on 250, 30, 4, and 0.5 mg of mouse kidney, the amounts of ds cDNA remaining on the beads reached 450, 67, 13, and 1.8 ng, respectively. Comparison of these data with those dealing with Sau3A I-released fragments (*2.3*) indicates that ~6 times lower cDNA amounts are recovered on beads that on Sau3A I-supernatants. The average size of Sau3A I-cut fragments is predicted to be 256 bp. The fraction that remains bound on the beads after Sau3A I digestion thus suggests that the average length of cDNA formed is ~1.5 kb, which seems quite reasonable.

[0054] The whole amount of BsmF I-released material is used for ditag formation, and we never attempted to quantify it. Nevertheless, the efficiency of BsmF I digestion can be checked when ≥ 4 mg of tissue is used for library generation. In this case, a Geiger counter allows to detect radioactivity in the BsmF I-supernatant.

4.2. Blunt ending of released cDNA tags

[0055] Different enzymes may be used for blunt ending BsmF I-released tags. In their original study, Velculescu *et al.* (7) carried out the blunt ending reaction with T4 DNA polymerase. In more recent applications, Klenow DNA polymerase was used (8) and is now recommended. It is also our experience that the success in library generation is very poor using T4 DNA polymerase. This likely comes from the fact that blunt ending with T4 DNA polymerase is carried out at 11°C (12). Such a low temperature allows protruding termini from unrelated cDNA tags to hybridize, and is thus expected to markedly decrease the amount of material available for the blunt ending reaction. We have successfully used Vent and sequencing grade T7 DNA polymerases to generate blunt ends. The procedure described in *protocol 5* involves T7 DNA polymerase.

Protocol 5. Release, blunt ending, and ligation of cDNA tags

[0056] *Equipment and reagents*

- BsmF I, 10X NEB IV buffer and 100X BSA.
- PCI.
- 10 M ammonium acetate.
- Sequencing grade T7 DNA polymerase (Pharmacia Biotech, ref. 27098503).
- 5X mix salt (200 mM Tris-HCl (pH 7.5), 100 mM $MgCl_2$, 250 mM NaCl).
- 2 mM dNTP.
- T4 DNA ligase (5 U/ µl) and 5X reaction buffer.

- 100% ethanol, 75% ethanol.
- Geiger counter.
- Water-baths equilibrated at 65°C, 42°C, and 16°C.

[0057] *Method*

1. Remove supernatant and immediately add on the beads 100 µl of the following mix: 87 µl H_2O, 10 µl 10X NEB IV, 1 µl 100X BSA, 2 µl BsmFI.
2. Incubate 2 h at 65°C. Vortex intermittently.
3. Chill 5 min at room temperature, collect the supernatant (which contains the ditags) and wash beads twice with 75 µl of ice-cold TE + BSA[a]. Pool all 3 supernatants (250 µl final volume) and add 60 µg glycogen to each of the 4 reaction tubes. Measure the radioactivity still present on the beads by Cerenkov counting (see text).
4. Add 250 µl (1 volume) PCI to all 4 supernatants.
5. Vortex, then centrifuge (10,000g) 10 min at 4°C. Transfer the upper (aqueous) phase to a fresh tube.
6. Precipitate with high ethanol concentration: add to the aqueous phase 125 µl 10 M ammonium acetate, 1.125 ml 100% ethanol, and centrifuge (15,000g) 20 min at 4°C.
7. Wash the pellet twice with 400 µl of 75% ethanol. Vacuum dry and resuspend the pellet in 10 µl LoTE.
8. Add 15 µl of 1X mix salt on each tube and heat 2 min at 42°C. Maintain tubes at 42°C and add 25 µl of the following mix: 7.5 µl H_2O, 5.5 µl 100 mM DTT, 11 µl dNTP mix, 1 µl T7 DNA polymerase. Incubate 10 min at 42°C.
9. Pool together tags ligated to hybrid(1A/1B) and hybrid(2A/2B). Rinse the tubes with 150 µl LoTE + 20 µg glycogen and add this solution to the pooled reactions (final volume: 250 µl). You have now 2 tubes (1 sample, 1 negative control).
10. Extract with equal volume of PCI and high concentration ethanol precipitate (see steps 4-6). Resuspend the pellet in 6 µl LoTE.
11. Ligate tags to form ditags by adding to the 6 µl-sample: 2 µl 5X mix ligase, 1 µl 10 mM ATP, and 1 µl of T4 DNA ligase (5 U/ µl). Proceed similarly for the negative control, incubate overnight at 16°C, then add 90 µl LoTE.

5. PCR Amplification

[0058] Considering the linkers and primers used in our studies, the desired PCR product is 110-bp long (90 bp of linkers derived sequences, and 20 bp of ditag).

5.1. PCR buffers and procedures

[0059] For PCR amplification of ditags, we use buffers and conditions different from those described by Velculescu et al.:

(a) amplification is performed with standard PCR buffers without DMSO and β-mercaptoethanol. Composition of our 10X PCR buffer is given *in protocol 3*. Promega buffer (ref. M1901) works equally well. The conditions used in our assay are as follows: 100 µM dNTP, 2.5 mM $MgCl_2$ (2 mM with Promega buffer), 0.5 µM primers, and 5 U *Taq* polymerase. High amounts of primers and *Taq* polymerase are used (standard reactions are generally performed with 0.1 µM primers and 1.25 U of enzyme) to ensure a high yield of ditags production. dNTP concentration is also slightly higher (100 vs. 50 µM) than for standard PCR amplifications. Very high dNTP concentrations should nevertheless be avoided since these are known to increase *Taq* polymerase-dependent misincorporations.

(b) as suggested initially (7), we still perform small scale PCR, purify the 110-bp fragment, then submit it to preparative PCR.

5.2. Number of PCR cycles

[0060] Before starting protocol 6, the optimal number of PCR cycles needs to be determined. This is best accomplished by performing duplicate PCR on 2% of the ligation product and sampling 7-µl aliquots at different cycles. The number of cycles will of course depend on the amount of starting material. For 250 mg tissue pieces, a PCR signal should be obtained with 18 cycles, and the plateau reached at 22-23 cycles. The 110-bp fragment should be largely predominant (amplified products of 90 and 100 bp are not unusual). Examples of PCR carried out on ditags generated from tiny amounts of cells (15,000 to 45,000) are given in Fig. 3. Using such low amounts of cells, the 110-bp product is no longer predominant. Nevertheless, if maximal yield is achieved with less than 30 cycles (as obtained from 45,000 cells in Fig. 3), a library which is fairly representative of the tissue can be generated. Small scale PCR (10 reactions,

[a]Final concentration of BSA: 0.1 mg/ ml.

step 1-5 in protocol 6) is performed on 2 µl and 4 µl aliquots of ligation product for macro and microamounts of tissue, respectively.

Protocol 6. PCR Amplification of Ditags

[0061] *Equipment and reagents*

- Automated thermal reactor (Hybaid).
- PCR reagents: *Taq* polymerase, Primers 1 and 2 at 20 pmol/ µl, 10X PCR buffer (see *protocol 3),* 1.25 mM dNTP, 100 mM $MgCl_2$, and 100 mM DTT.
- β-Agarase and 10X reaction buffer (New England Biolabs, ref. 392L).
- Sau3A I reaction buffer and 100X BSA.
- Low melting point (LMP) agarose (Life Technologies, ref. 15517-022).
- TBE 10X (1.12 M Tris, 1.12 M boric acid, 20 mM EDTA)
- Vertical gel electrophoresis unit, with 20X20 cm plates, 1.5mm thick spacers, and preparative comb.
- 10-bp DNA ladder.
- Bromophenol blue loading buffer (0.125% bromophenol blue, 10% ficoll 400, 12.5 mM EDTA) filtered on 0.45-µm membrane.

[0062] *Method*

1. Prepare a master mix with the following reagents : 5 µl 10X PCR Buffer, 1 µl 100 mM $MgCl_2$, 8 µl 1.25 mM dNTP, 2.5 µl primer 1, 2.5 µl primer 2, 27 µl H_2O (multiply these quantities by the number of reactions tubes (usually 12)). Dispense equal aliquots (46 µl) into PCR tubes and add 4 µl of DNA sample (10 tubes), negative control, or H_2O.
2. Transfer the tubes in the thermal reactor and heat 2 min at 80°C (hot start conditions).
3. Add in each tube 50 µl of the following mix: 5 µl 10X PCR Buffer, 4 µl 100 mM DTT, 40 µl H_2O, 1 µl *Taq* Polymerase. Add a drop of mineral oil if necessary according to your thermal cycler.
4. Perform PCR at the following temperatures : 30 sec at 94°C, 30 sec at 58°C, and 45 sec at 70°C (27-30 cycles), followed by one cycle with an elongation time of 5 min.
5. Analyse an aliquot of each tube (7 µl) on a 3% agarose gel using 10-bp DNA ladder as marker.
6. If yield is satisfactory (see Fig. 3), pool the 10 PCR tubes in two 1.5 ml tubes. Add 30 µg glycogen in each tube, extract with PCI. Recover the aqueous phase and precipitate the DNA by centrifugation after adding 0.1 volume 3 M sodium acetate and 2.5 volumes 100% ethanol. Wash the pellet with 75% ethanol, vacuum dry, and resuspend in 300 µl LoTE. Add 75 µl of bromophenol blue loading buffer.
7. Electrophorese the PCR product through a 3% LMP agarose vertical gel (warm plates 15 min at 55°C before pouring the gel). Run until bromophenol blue has reached bottom of the gel (~3 h).
8. Cut out the 110-bp fragment from gel and place agarose slice in a 2 ml tube. Add 0.1 volume 10X 13-agarase mix, heat 10 min at 70°C, then 10 min at 40°C, and add β-Agarase (6 U/ 0.2g of agarose). Incubate 1h30 at 40°C. Add 30 µg glycogen. Extract with PCI and ethanol precipitate as indicated in section 6. Resuspend the pellet in 300 µl LoTE.
9. After determination of the optimal number of PCR cycles (usually 12), perform large scale PCR (140-150 reactions) using 2 µl of DNA and the protocol described in sections 1-4.
10. Pool PCR reactions in 2 ml tubes. Extract with PCI, ethanol precipitate (section 6) and wash the pellet twice with 75% ethanol. Resuspend the dry pellets in a final volume of 470 µl 1X mix Sau3A I.

5.3. Purification and reamplification

[0063] The 110-bp PCR product can be purified either on a 12% polyacrylamide (7) or a 3% agarose slab gel. To avoid overloading and achieve efficient purification, pool no more than 10-12 PCR reactions on an agarose gel and slice agarose as close as possible to the 110-bp fragment. Purification and optimal number of PCR cycles should then be tested on duplicate 2 µl aliquots of the purified product. A single band of 110 bp should now be obtained. The absence of interference from other amplified products is essential to produce large amounts of the 110-bp fragment.

6. Ditags isolation, concatenation, and cloning of concatemers

6.1. Ditags isolation

[0064] Two important points need to be addressed for ditags purification. First, since the total mass of linkers is nearly

five times that of ditags, a highly resolutive polyacrylamide gel is required to thoroughly purify ditags. Second, the short length of ditags makes them difficult to detect on gel by ethidium bromide staining. This problem can be overcome by staining the gel with SYBR Green I (or equivalent products) which ensures a lower detection threshold than ethidium bromide (0.1 instead of 2 ng DNA). To obtain high sensitivity, loading buffer should not contain bromophenol blue (bromophenol blue comigrates with ditags). The gel is stained after migration in a polypropylene or PVC container.

[0065] Ditags do not run as a single band on polyacrylamide gel. This may come from subtle effects of base composition on electrophoretic mobility and/ or some wobble for BsmF I digestion (7). We cut out from gel all the material ranging from 22 to 26 bp. The elution procedure is labour intensive but provides ditags that can be concatenated efficiently.

6.2. Concatenation

[0066] Starting from 150 PCR reactions, at least 1 µg ditags should be obtained. The optimal ligation time depends on the amount of ditags and on the purity of the preparation. We usually perform ligation for 2 h. When yield is high (≥ 1.4 µg), we set up two ligation reactions, and allow them to proceed for 1 or 2 h. The corresponding concatemers are then separately purified on a 8% agarose gel.

Protocol 7. Ditags isolation and concatenation

[0067] *Equipment and Reagents*

- Sau3A I, 10X reaction buffer and 100X BSA.
- 50X TAE (2 M Tris, 57% glacial acetic acid, 50 mM EDTA)
- 12% polyacrylamide gel : 53.6 ml H_2O, 24 ml 40% acrylamide (19:1 acrylamide:bis), 1.6 ml 50X TAE, 800 µl 10% ammonium persulfate, 69 µl TEMED.
- 10-bp DNA ladder.
- SYBR Green I stain (FMC Bioproducts, ref. 50513).
- T4 DNA Ligase 5 U/ µl and 5X ligation mix; 10 mM ATP.
- Vertical gel electrophoresis unit, with 20X20 cm plates, 1.5mm thick spacers, and preparative comb.
- Xylene cyanole loading buffer (0.125% xylene cyanole, 10% ficoll 400, 12.5 mM EDTA).
- Spin X microcentrifuge tubes (Costar, ref. 8160)

[0068] *Method*

1. Save 1 µl of the 110-bp DNA fragment *(section 10 of protocol 6)* and digest the remaining by adding 5 µl 100X BSA and 25 µl Sau3A I. Incubate overnight at 37°C in hot-air incubator.
2. Check for Sau3A I digestion: analyse 1 µl of uncut DNA, 1 µl and 3 µl of Sau3A I digestion (use bromophenol blue loading buffer and xylene cyanole loading buffer for uncut and Sau3A I-digested DNA, respectively) on a 3% agarose gel. Most (>80%) of the 110-bp fragment has been digested, and a faint band, corresponding to the ditags can now be detected at ~25 bp.
3. Add 125 µl xylene cyanole loading buffer to the digested DNA sample and load on a preparative 12% polyacrylamide vertical gel in 1X TAE. Run at 30mA until bromophenol blue of the size marker is 12 cm away from the well.
4. Transfer the gel in SYBR Green I stain at 1:10,000 dilution in 1X TAE. Wrap the container in aluminium foil and stain gel for 20 min. Visualise on UV box.
5. Cut out the ditags band (24-26 bp) and transfer acrylamide slices in 0.5 ml tubes (for a 20-cm wide gel use 8 tubes). Pierce the bottom of 0.5 ml tubes with a 18-gauge needle. Place the tubes in 2 ml tubes and spin 5 min at 10,000 g. Prepare the following elution buffer for each tube : 475 µl LoTE, 25 µl 10 M ammonium acetate, 5 µg glycogen. Add 250 µl elution buffer in each 0.5 ml tube and centrifuge again. Discard 0.5 ml tubes and add 250 µl elution buffer directly in each 2 ml tube. Incubate overnight at 37°C in hot-air incubator.
6. Prepare a series of 16 SpinX microcentrifuge tubes: add 20 µg glycogen in each collection tube. Transfer content of each 2 ml tube (~600 µl) to two SpinX microcentrifuge tubes. Spin 5 min at 13,000 g. Transfer 350 µl of eluted solution into 1.5 ml tubes (10-11 tubes), extract with PCI, perform high concentration ethanol precipitation. Wash twice with 75% ethanol, vacuum dry, and pool all pellets in 15 µl LoTE.
7. Measure the amount of purified ditags by dot quantitation (12) using 1 µl of sample. Total DNA at this stage is usually 1 µg, but a library can still be generated with 400ng.
8. Ligate ditags to form concatemers: add to your sample (14 µl) 4.4 µl 5X mix ligase, 2.2 µl 10 mM ATP, and 2.2 µl concentrated (5 U/ µl) T4 DNA ligase.
9. Incubate 2h at 16°C. Stop the reaction by adding 5 µl of bromophenol blue loading buffer and store at -20°C.

6.3. Purification of concatemers

[0069] Concatemers are heated at 45°C for 5 min immediately before loading on gel to separate unligated cohesive ends. Concatemers form a smear on the gel from about 100 bp to several kbp (Fig. 4) and can be easily detected using SYBR Green I stain. All fragments >300 bp (i.e. with 25 or more tags) are potentially interesting for library construction. We usually cut out fragments of 350-600, 600-2000, and >2000 bp and generate a first library using 600-2000 bp DNA fragments. Longer fragments will be more informative but are expected to be cloned with poor efficiency.

Protocol 8. Purification and cloning of concatemers

[0070] *Equipment and reagents.*

- SYBR Green I stain
- Vertical gel electrophoresis unit, with 20X20 cm plates, 1.5mm thick spacers, and 20-well comb.
- 8% polyacrylamide gel : 61.6 ml H_2O, 16 ml acrylamide 40% (37.5:1 acrylamide:bis), 1.6 ml 50X TAE, 800 µl 10% ammonium persulfate, 69 µl TEMED.
- 100-bp DNA ladder (Life Technologies, ref. 15628-019)
- T_4 DNA ligase 1 U/ µl (Life Technologies, ref. 15224-017) and 5X reaction buffer.
- 10 mM ATP.
- pBluescript II, linearized with BamH I and dephosphorylated.
- *E. coli* XL2 Blue ultracompetent cells (Stratagene, ref. 200150).

[0071] *Method*

1. Heat sample 5 min at 45°C and load into one lane of a 20 wells 8% acrylamide gel. Run at 30 mA until bromophenol blue is 10-12 cm from the well.
2. Stain the gel with SYBR Green I as described in *protocol 7* and visualise on UV box.
3. Concatemers form a smear on gel with a range from about 100 bp to the gel well (Fig. 4).Cut out regions containing DNA of 350-600, 600-2000, and >2000 bp. Purify separately DNA of each three slices as described in section 5-6 of protocol 7 (a 1-h incubation period of gel slices in LoTE/ammonium acetate solution is sufficient). Resuspend the pellet in 6 µl LoTE and generate a first library using concatemers of 600-2000 bp.
4. Mix 6 µl of concatemers and 2 µl (25 ng) of BamH I-cut pBluescript II. Heat 5 min at 45°C then chill on ice.
5. Add 3 µl 5X mix ligase, 1 µl H_2O, 1.5 µl 10 mM ATP, and 1.5 µl T4 DNA ligase (1 U/ µl). Mix and incubate overnight at 16°C.
6. Add 20 µg glycogen and 285 µl LoTE and extract with PCI. Ethanol precipitate, wash twice with 75% ethanol, vacuum dry, and resuspend the pellet in 12 µl LoTE.
7. Transform *E. coli* XL2 Blue ultracompetent cells with 1/3 (4 µl) of ligation reaction according to the manufacturer's instructions. Plate different volumes (5 µl, 10 µl, 20 µl, 40 µl) of transformation mix onto Petri dishes containing Luria agar supplemented with ampicillin, X-gal, and IPTG. Incubate 15-16hr at 37°C. Save the remaining (~900 µl) transformation solution (add 225 µl 80% glycerol, mix intermittently for 5 min, and store at -80°C). It will be used to plate additional bacteria if library appears correct.
8. Count insert-free (*i.e.* blue) and recombinant (*i.e.* white) bacterial colonies on each plate. The fraction of recombinant colonies should be >50%, and their total number should be in the range of 10,000-60,000 for 1 ml of transformation mix.

6.4. Cloning of concatemers

[0072] Concatemers can be cloned and sequenced in a vector of choice. We currently clone concatemers in pBluescript II linearized with BamH I and dephosphorylated by calf intestinal alkaline phosphatase treatment. Any kind of vector with a BamH I site in the multiple cloning site will be suitable. Velculescu et al.(7, 8) use pZero-1 from Invitrogen which only allows recombinants to grow (DNA insertion into the multiple cloning site disrupts a lethal gene). The competent cells and transformation procedures (heat shock or electroporation) can also be changed according to your facilities. Whatever your choice, it is important to use bacterial cells allowing very high cloning efficiency (≥ $5X10^9$ transformants/ µg of supercoiled DNA). An important point is to evaluate the number of clones *(protocol 8, step 8)* obtained in the library. Since a large number (1,000-2,000) of clones will be sequenced, the total number of recombinants should be > 10,000.

[0073] Library screening can be performed by PCR or DNA miniprep. In our hands, DNA miniprep provides more reproducible amounts of DNA than PCR, and avoids false positive signals. We use Qiaprep 8 miniprep kit (Qiagen,

ref. 27144) which enables to perform 96 minipreps in ~2 hours. Plasmid DNA is eluted from Qiagen columns with 100 µl of elution buffer; 5 µl are digested to evaluate insert size, and if insert is > 200 bp, 5 µl are directly used for DNA sequencing.

7. Data analysis

7.1. Software

[0074] Once cloned concatemers have been sequenced, tags must be extracted, quantified, and identified through possible data bank matches. Two softwares have been written to reach these goals.

[0075] SAGE software (7) was written in Visual Basic and is operating on personal computers through the Microsoft Windows system. It extracts tags from text sequence files, quantify them, allows to compare several libraries, and provides links to GenBank data downloaded from CD-ROM flat files or over the Internet. The latter function enables rapid identification of tags originating from characterised genes or cDNAs. However, description of EST sequences is truncated, which constrains to look for individual GenBank reports. SAGE software also includes several simulating tools which allows, for example, to assess the significance of differences observed between two libraries, and to evaluate the sequencing accuracy.

[0076] The second software is currently developed at the University of Montpellier-2 (France) by J. Marti and co-workers, as part of a database (CbC, for Cell by Cell) intended to store and retrieve data from SAGE experiments. Scripts for extraction of data are developed in C language under Unix environment and the database management system implemented in Acces®. Text files are concatenated to yield the working file from which tag sequences are extracted and enumerated. Treatment of raw data involves identification of vector contaminants, truncated and repeated ditags (see below). For experiments on human, mouse and rat cell samples, the tags are searched in the non-redundant set of sequences provided by the UniGene collection. These data can be loaded from the anonymous FTP site: ncbi.nlm.gov/repository/unigene/. Useful files are Hs.data.Z and Hs.seq.uniq.Z for man, and similar files for mouse and rat. The results are displayed in a table which provides the sequence of each tag, its number of occurrences with the matching cluster number (Hs# for *homo sapiens*, Mm# for *Mus musculus* and Rn# for *Rattus norvegicus)*, and other data extracted from the source files, including GenBank accession numbers. For human genes, when available, a link is automatically established with GeneCards, (http://bioinfo.weizmann.ac.il/cards/) allowing to get additional information.

7.2. Library validity

7.2.1. Inserts length

[0077] Initial assessment of library quality will be obtained from screening for inserts length. A good library should contain >60% clones with inserts ≥ 240 bp (20 tags). Only one DNA strand is sequenced, since accuracy is obtained from the number of tags recorded, rather than from the quality of individual runs. Depending on your budget and sequencing facilities, either all clones or only the most informative ones will be sequenced. It should be noted that the average length of inserts does not fit with that of gel-purified concatemers. Although we usually extract 600-2000 bp long concatemers, most of the clones have inserts <600 bp, and we never get inserts >800 bp. A number of reasons can explain such a paradoxical result. Indeed, long inserts are known to be cloned with poor efficiency. In addition, they are expected to contain several repeats or inverted repeats, and may thus form unstable plasmid constructs. Supporting this interpretation, it has been demonstrated (14), and we have also observed, that efficient removal of linkers (which represents up to 20% of total tags in poor libraries) increases the average length of cloned inserts. At any rate, it is worth to emphasise that similar biological information is obtained from libraries with short and long inserts.

7.2.2. Gene expression pattern

[0078] The basic pattern of gene expression in eukaryotic cells have been established long ago by kinetics analysis of mRNA-cDNA hybridization (15, 16). In a "typical" mammalian cell, the total RNA mass consists of 300,000 molecules, corresponding to ~12,000 transcripts which divide into three abundance classes. A very small number of mRNAs (~10) are expressed to exceedingly high levels (3,000-15,000 copies/ cell). A larger number of mRNAs (~500) reaches an expression level in the range of 100-500 copies/ cell. Finally, the majority of mRNAs (>10,000) are poorly expressed (10-100 copies/ cell). This basic pattern should be observed in SAGE or SADE libraries. However, before translating tags abundance in a definite gene expression profile, the data must be scrutinised for artefacts encountered in library construction.

7.2.3. Occurrence of linker-derived sequences

[0079] As mentioned above, some libraries display a high amount of linker sequences. If this amount is 20% or more, sequencing will be quite expensive, and it is better to start again from the RNA sample. Library contamination with 10-15% of linker sequences is acceptable, 5-10% is good, and < 5% is excellent. In addition to the two perfect linker matches (GTCCCTGTGC, and GTCCCTTCCG), reading ambiguities can lead to sequences with one mismatch. These linker-like sequences are also easily identified since, assuming efficient enzymatic cleavage, the probability of having adjacent Sau3A I and BsmF I sites in the concatemers is normally zero. Linker and linker-like sequences can be automatically discarded using SAGE or CbC software, and their relative amounts can be used to evaluate the sequencing accuracy (see *7. 1.*).

7.2.4. Duplicate ditags

[0080] Another category of sequences that must be deleted are those corresponding to duplicate ditags. Indeed, except for peculiar tissues (e.g. lactating gland or laying hen oviduct) in which one or a very small number of transcripts constitutes the bulk of the mRNA mass, the probability for any two tags to be found several times in the same ditag is very small. Elimination of repeated ditags will therefore correct for preferential PCR amplification of some targets, and for picking several bacterial colonies originating from the same clone. Most ditags (>95%) generally occur only once when the library is constructed from macroamounts of tissue. For microlibraries, the percent of unique ditags is generally lower. When it is no longer compatible (<75%) with efficient data acquisition, it is recommended to start again from the first (small scale) PCR (see *protocol 6).* Duplicate ditags are automatically retrieved from the sequence files by SAGE and CbC softwares.

7.3. Number of tags to be sequenced

[0081] The number of tags to be analysed will obviously depend on the application and tissue source. As a matter of fact, reducing the tissue complexity through isolation of defined cell populations will allow to markedly diminish the minimum number of tags for accurate analysis, and to better correlate molecular and physiological phenotypes.

[0082] Delineation of the most expressed genes (>500 copies/ cell) in one tissue, and comparison with their expression level in another one, will require to sequence only a few thousand tags (Fig. 5). Analysing 5,000 tags, 300 will be detected at least 3 times. Since automated sequencers can read 48-96 templates simultaneously, 10,000 tags will be recorded from 5-10 gels if the average number of tags/ clone is ~20.

[0083] The most difficult projects are those aiming to compare gene expression profiles in the same tissue under two physiological or pathological conditions. Differentially expressed genes could belong to any of the three abundance classes and, furthermore, they can be either up- or down-regulated. A reasonable number of tags to be sequenced would be in the range of 30,000-50,000. The probability (P) of detecting a sequence of a given abundance can be calculated from the Clarke and Carbon (17) equation (N=ln(1-P)/ln(1-x/n)), where N is the number of sequence analysed, x is the expression level, and n is the total number of mRNAs per cell (~300,000). Thus, the analysis of 30,000 and 50,000 tags will provide a 95% confidence level of detecting transcripts expressed at 30 and 18 copies/ cell, respectively. Most up-regulation processes will be therefore assessed. For example, tags corresponding to poorly expressed transcripts may be detected 1 and ≥ 5 times in control and experimental conditions, respectively. However, we have to be aware that the possibility of assessing down-regulation processes will be less exhaustive. It will only concern tags present > 5-10 times in the control condition, which excludes from the analysis part of the poorly expressed transcripts.

[0084] In the here above Table I, which corresponds to the characterisation of the most abundant nuclear transcripts in the mouse outer medullary collecting duct (OMCD) and establishes their differential expression in the medullary thick ascending limb (MTAL), the two left columns correspond to the data illustrated in Figure 5, and provide the abundance of each tag in the two libraries. The third column provides the sequence of the tags. The right column indicates results of individual BLAST search in GenBank, carried out using a 14-bp sequence (the Sau3A I recognition sequence, plus the 10 bp specific for each tag).

References

[0085]

1. DeRisi, J.L., Vishwanath, R.Y., and Brown, P.O. (1997). *Science,* **278**, 680.
2. Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., and Lockhart, D.J. (1997). *Nature* Biotechnology, **15**, 1359.
3. Gress, T.M., Hoheisel, J.D., Lennon, G.G., Zehetner, G., and Lehrach, H (1992). *Mamm. Genome,* **3**, 609.

4. Piétu, G., Alibert, A., Guichard, V., Lamy, B., Bois, F., et al. (1996). *Genome Research,* **6**, 492.
5. Adams, M.A., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., et al. (1995). *Nature,* **377 (No. 6547S)**, 3.
6. Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., et al. (1992). *Nature Genet.,* **2**, 173.
7. Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. (1995). *Science,* **270**, 484.
8. Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M., et al. (1997). *Cell,* **88**, 243.
9. Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., et al. (1997). *Science,* **276**, 1268.
10. Polyak, K., Xia, Y., Zweier, J.L., Kinzler, K.W., and Vogelstein, B. (1997). *Nature,* **389**, 300.
11. Sade, D.A.F. (1990). Oeuvres complètes. Gallimard, Paris.
12. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., et al. (1993, updated quaterly). *Current protocols in Molecular Biology.* Greene Publishing Associates and Wiley-Interscience, New York.
13. Chomczynski, P., and Sacchi, N. (1987). *Anal. Biochem.,* **162,** 156.
14. Powell, J. (1998). *Nucleic Acids Res.,* **26**, 3445.
15. Hastie, N.D., and Bishop, J.O. (1976) *Cell,* **9,** 761.
16. Hereford, L.M., and Rosbash, M. (1977). *Cell,* **10**,453.
17. Clarke, L., and Carbon, J. (1976). *Cell,* **9**, 91.

[0086] As evident from the above, the invention is not at all limited to its embodiment, implementation and application which have just been defined more explicitly; it embraces, on the contrary, all the variants which may occur to a specialist in this field, without departing from the framework or the scope of the present invention.

Annex to the description

[0087]

SEQUENCE LISTING

<110> COMMISSARAIT A L'ENERGIE ATOMIQUE
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

<120> MICROASSAY FOR SERIAL ANALYSIS OF GENE EXPRESSION AND APPLICATIONS THEREOF.

<130> BLOcp263EP51

<140>
<141>

<160> 27

<170> PatentIn Ver. 2.1

<210> 1
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:LINKER

<400> 1
gatcgtccc 9

<210> 2
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:LINKER 1A

<400> 2
ttttgccagg tcactcaagt cggtcattca tgtcagcaca gggac 45

<210> 3
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:LINKER 1B

<400> 3
gatcgtccct gtgctgacat gaatgaccga cttgagtgac ctggca 46

<210> 4
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:LINKER2A

<400> 4
tttttgctca ggctcaaggc tcgtctaatc acagtcggaa gggac 45

<210> 5
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:LINKER 2B

<400> 5
gatcgtccct tccgactgtg attagacgag ccttgagcct gagcaa 46

<210> 6
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PRIMER

<400> 6
gccaggtcac tcaagtcggt catt 24

<210> 7
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:PRIMER

<400> 7
tgctcaggct caaggctcgt cta 23

<210> 8
<211> 10
<212> DNA
<213> Mus sp.

<400> 8
gtggcagtgg 10

<210> 9
<211> 10
<212> DNA
<213> Mus sp.

<400> 9
ttataatttg 10

<210> 10
<211> 10
<212> DNA
<213> Mus sp.

<400> 10
tggcagtggg 10

<210> 11
<211> 10
<212> DNA
<213> Mus sp.

<400> 11
tgactccctc 10

<210> 12
<211> 10
<212> DNA
<213> Mus sp.

<400> 12
aagtttaaat 10

<210> 13
<211> 10
<212> DNA
<213> Mus sp.

<400> 13
agcaagcagg 10

<210> 14
<211> 10
<212> DNA
<213> Mus sp.

<400> 14
caaaaagcta 10

<210> 15
<211> 10
<212> DNA
<213> Mus sp.

<400> 15
acattcctta 10

<210> 16
<211> 10
<212> DNA
<213> Mus sp.

<400> 16
accgaccgca 10

<210> 17
<211> 10
<212> DNA
<213> Mus sp.

<400> 17
cagaagaagt 10

<210> 18
<211> 10
<212> DNA
<213> Mus sp.

<400> 18
aaataaagtt 10

<210> 19
<211> 10
<212> DNA
<213> Mus sp.

<400> 19
agaagcagtg 10

<210> 20
<211> 10
<212> DNA
<213> Mus sp.

<400> 20
tgatgccctc 10

<210> 21
<211> 10
<212> DNA
<213> Mus sp.

<400> 21
aggctactac 10

<210> 22
<211> 10
<212> DNA
<213> Mus sp.

<400> 22
gctcattgga 10

<210> 23
<211> 10
<212> DNA
<213> Mus sp.

<400> 23
gctttcagca 10

<210> 24
<211> 10
<212> DNA
<213> Mus sp.

<400> 24
gtgactgggt 10

<210> 25
<211> 10
<212> DNA
<213> Mus sp.

<400> 25
tgaccaaggc 10

<210> 26
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:linker

<400> 26
gtccctgtgc 10

<210> 27
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:linker

<400> 27
gtcccttccg 10

Claims

1. Method of obtaining a library of tags able to define a specific state of a biological sample, characterised in that it comprises the following successive steps:

(1) extracting in a single-step mRNA from a small amount of a biological sample using oligo(dT)$_{25}$ covalently bound to paramagnetic beads,
(2) generating a double strand cDNA library, from said mRNA according to the following steps:

* synthesising the 1st strand of said cDNA by reverse transcription of said mRNA template into a 1st complementary single-strand cDNA, using a reverse transcriptase lacking Rnase H activity,
* synthesising the 2nd strand of said cDNA by nick translation of the mRNA, in the mRNA-cDNA hybrid form by an *E. coli* DNA polymerase,

(3) cleaving the obtained cDNAs using the restriction endonuclease Sau3A I as anchoring enzyme,
(4) separating the cleaved cDNAs in two aliquots,
(5) ligating the cDNA contained in each of said two aliquots via said Sau3A I restriction site to a linker consisting of one double-strand cDNA molecule having one of the following formulas:

$$GATCGTCCC\text{-}X_1 \text{ or } GATCGTCCC\text{-}X_2,$$

wherein X_1 and X_2, which comprise 30-37 nucleotides and are different, include a 20-25 bp PCR priming site with a Tm of 55°C-65°C, and
wherein GATCGTCCC correspond to a Sau3A I restriction site joined to a BsmF I restriction site,

(6) digesting the products obtained in step (5) with the tagging enzyme BsmF I and releasing linkers with anchored short piece of cDNA corresponding to a transcript-specific tag, said digestion generating BsmF I tags specific of the initial mRNA,
(7) blunt-ending said BsmF I tags with a DNA polymerase, preferably T7 DNA polymerase or Vent polymerase and mixing the tags ligated with the different linkers,
(8) ligating the tags obtained in step (7) to form ditags with a DNA ligase,
(9) amplifying the ditags obtained in step (8) with primers comprising 20-25 bp and having a Tm of 55°-65°C,
(10) isolating the ditags having between 20 and 28 bp from the amplification products obtained in step (9) by digesting said amplification products with the anchoring enzyme Sau3A I and separating the digested products on an appropriate gel electrophoresis,
(11) ligating the ditags obtained in step (10) to form concatemers, purifiying said concatemers and separating the concatemers having more than 300 bp,
(12) cloning and sequencing said concatemers and
(13) analysing the different obtained tags.

2. Method according to claim 1, characterised in that in step (2), said synthesis of the 1st strand of said cDNA is performed with Moloney Murine Leukaemia Virus reverse transcriptase (M-MLV RT), and oligo(dT)$_{25}$ as primers.

3. Method according to claim 1 or to claim 2, characterised in that the linkers of step (5) are preferably hybrid DNA molecules formed from linkers 1A and 1B or from linkers 2A and 2B, having the following formulas:

linker 1A: 5'-TTTTGCCAGGTCACTCAAGTCGGTCATTCATGTCAGCACAGGGAC-3'

linker 1B: 5'-GATCGTCCCTGTGCTGACATGAATGACCGACTTGAGTGACCTGGCA-3'

or

linker 2A: 5'-TTTTTGCTCAGGCTCAAGGCTCGTCTAATCACAGTCGGAAGGGAC-3'

linker 2B: 5'-GATCGTCCCTTCCGACTGTGATTAGACGAGCCTTGAGCCTGAGCAA-3'.

4. Method according to claims 1 to 3, characterised in that the amount of each linker in step (5) is at most of 8-10 pmol and preferably comprised between 0.5 pmol and 8 pmol for initial amounts of respectively 10-40 ng of mRNAs and 5 µg of mRNAs.

5. Method according to claims 1 to 4, characterised in that the primers of step (9) have preferably the following formulas:

5'-GCCAGGTCACTCAAGTCGGTCATT-3'

5'-TGCTCAGGCTCAAGGCTCGTCTA-3'.

6. Method according to claims 1 to 5, characterised in that the biological sample of step (1) preferably, comprises ≤ 5.10^6 cells, corresponding to at most 50 µg of total RNA or 1 µg of poly(A) RNA.

7. Method according to claims 1 to 6, characterised in that said tissue sample is from kidney, more specifically from nephron segments corresponding to about 15,000 to 45,000 cells, corresponding to 0.15-0.45 µg of total RNA.

8. Use of a library of tags obtained according to the method of claims 1 to 7, for assessing the state of a biological sample.

9. Use of the tags obtained according to claims 1 to 7 as probes.

10. Method of determination of a gene expression profile, characterised in that it comprises :

- performing steps (1) to (13) according to claim 1 and
- translating cDNA tag abundance in gene expression profile.

11. Method according to claim 10, characterised in that the gene expression profile obtained in mouse outer medullary collecting duct (OMCD) and in mouse medullary thick ascending limb (MTAL) is as specified in Table I.

12. A kit useful for detection of gene expression profile, characterised in that the presence of a cDNA tag obtained from the mRNA extracted from a biological sample, is indicative of expression of a gene having said tag sequence at an appropriate position, i.e. immediately adjacent to the most 3' Sau3A I site in said cDNA, the kit comprising further to usual buffers for cDNA synthesis, restriction enzyme digestion, ligation and amplification,

- containers containing linker consisting of one double-strand cDNA molecule having one of the following formulas:

GATCGTCCC-X_1 or GATCGTCCC-X_2,

wherein X_1 and X_2, which comprise 30-37 nucleotides and are different, include a 20-25 bp PCR priming site with a Tm of 55°C-65°C, and
wherein GATCGTCCC correspond to a Sau3A I restriction site joined to a BsmF I restriction site, and

- containers containing primers comprising 20-25 bp and having a Tm of 55°-65°C

13. Kit according to claim 12, characterised in that it preferably contains:

- containers containing hybrid DNA molecules formed from linkers 1A and 1B or from linkers 2A and 2B, having the following formulas:

linker 1A: 5'-TTTTGCCAGGTCACTCAAGTCGGTCATTCATGTCAGCACAGGGAC-3'

linker 1B: 5'-GATCGTCCCTGTGCTGACATGAATGACCGACTTGAGTGACCTGGCA-3'

or

linker 2A: 5'-TTTTTGCTCAGGCTCAAGGCTCGTCTAATCACAGTCGGAAGGGAC-3'

linker 2B: 5'-GATCGTCCCTTCCGACTGTGATTAGACGAGCCTTGAGCCTGAGCAA-3',

and

- containers containing the following primers:

5'-GCCAGGTCACTCAAGTCGGTCATT-3'

5'-TGCTCAGGCTCAAGGCTCGTCTA-3'.

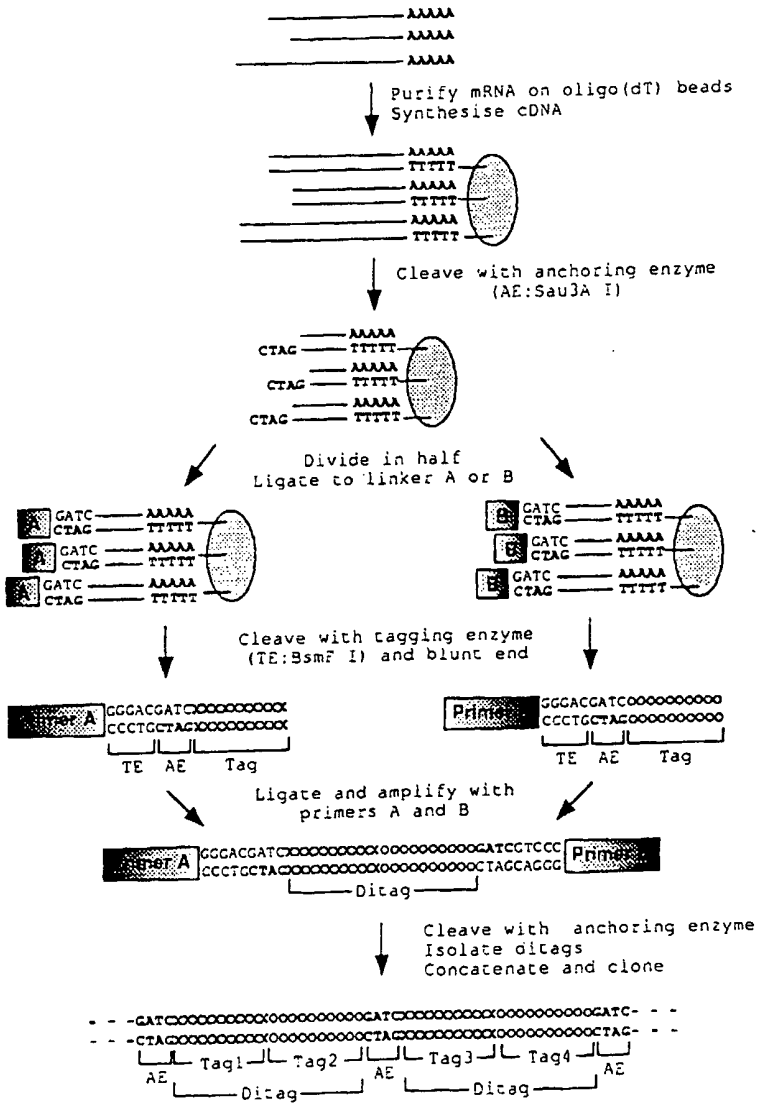

Purify mRNA on oligo(dT) beads
Synthesise cDNA
Cleave with anchoring enzyme
(AE:Sau3A I)
Divide in half
Ligate to linker A or B
Cleave with tagging enzyme
(TE:BsmF I) and blunt end
TE AE Tag
Ligate and amplify with
primers A and B
Ditag
Cleave with anchoring enzyme
Isolate ditags
Concatenate and clone
Tag1
Tag2
Tag3
Tag4
AE
Ditag

FIGURE 1

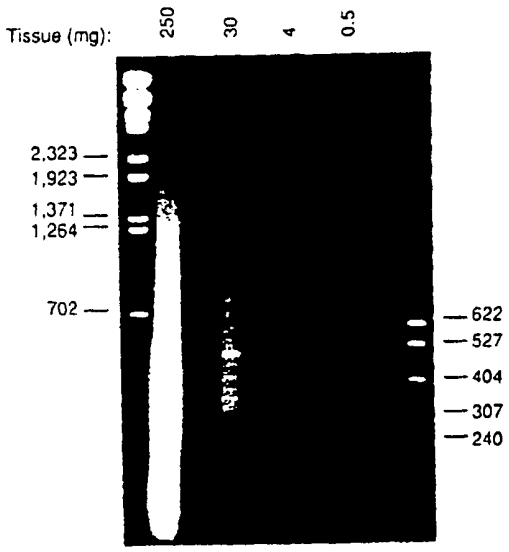

Tissue (mg):
250
30
4
0.5
2,323
1,923
1,371
1,264
702
622
527
404
307
240

FIGURE 2

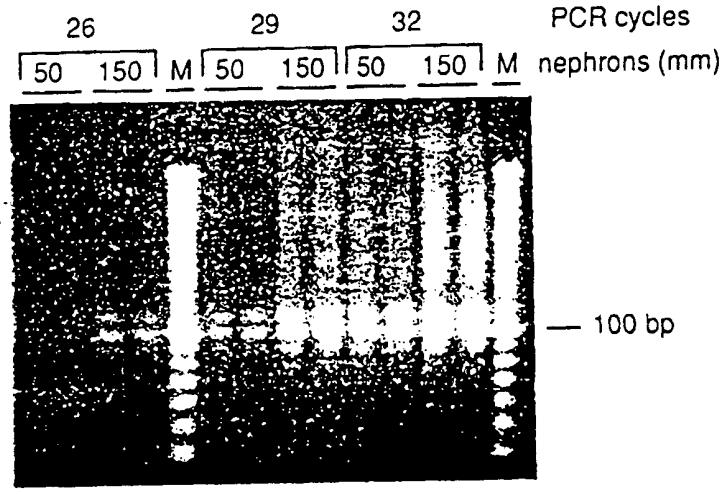


FIGURE 3

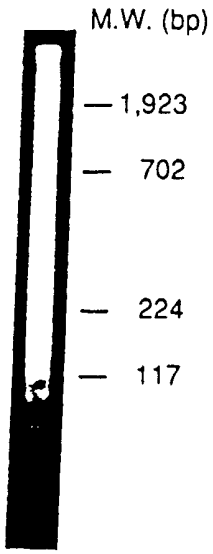


FIGURE 4

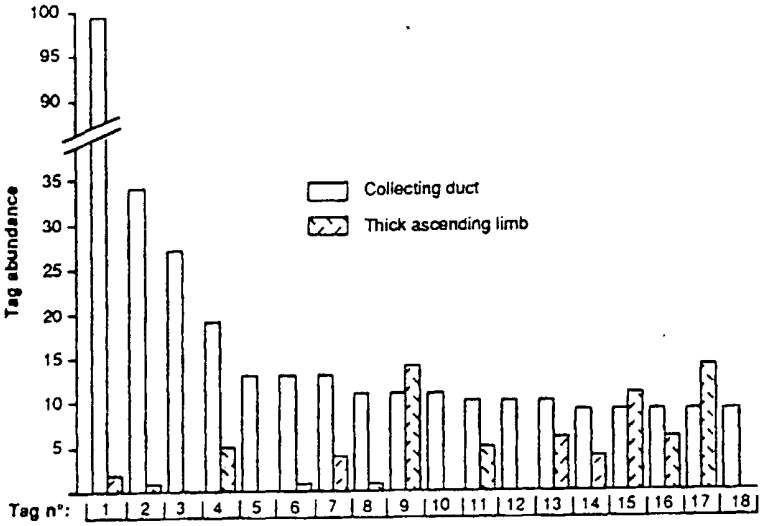

100
95
90
35
30
25
20
15
10
5
Tag abundance
Collecting duct
Thick ascending limb
Tag n°:
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

FIGURE 5

EP 1 024 201 A1

European Patent Office

PARTIAL EUROPEAN SEARCH REPORT

which under Rule 45 of the European Patent Convention shall be considered, for the purposes of subsequent proceedings, as the European search report

Application Number EP 99 40 0189

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Y	EP 0 761 822 A (THE JOHN HOPKINS UNIVERSITY SCHOOL OF MEDICINE) 12 March 1997 (1997-03-12) * the whole document * ---	1-8, 10-13	C12Q1/68
Y	AUGENSTEIN S.: "Superparamagnetic beads: applications of solid-phase RT-PCR" AMERICAN BIOTECHNOL. LABOR. (ISSN 0749-3223), vol. 12, no. 6, May 1994 (1994-05), pages 12-14, XP002116023 US * the whole document * ---	1	
Y	WO 97 29211 A (THE GOVERNMENT OF THE UNITED STATES OF AMERICA) 14 August 1997 (1997-08-14) * page 4, line 10 - page 6, line 2 * * page 8, line 32 - page 9, line 29; claims; figure 1 * -----	1-8, 10-13	
			TECHNICAL FIELDS SEARCHED (Int.Cl.7) C12Q

INCOMPLETE SEARCH

The Search Division considers that the present application, or one or more of its claims, does/do not comply with the EPC to such an extent that a meaningful search into the state of the art cannot be carried out, or can only be carried out partially, for these claims.

Claims searched completely :

Claims searched incompletely :

Claims not searched :

Reason for the limitation of the search:

see sheet C

Place of search	Date of completion of the search	Examiner
MUNICH	21 September 1999	Luzzatto, E

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C07)

European Patent Office

INCOMPLETE SEARCH SHEET C

Application Number
EP 99 40 0189

Claim(s) searched completely:
1-8,10-13

Claim(s) not searched:
9

Reason for the limitation of the search:

Claim 9 relates to the use of gene-specific sequences ("tags") as probes, without however relating to any technical feature of the said sequences since the method of obtention as such does not technically characterise them to any extent. No search was therefore possible with respect to the subject-matter of claim 9.

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 99 40 0189

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

21-09-1999

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 761822 A	12-03-1997	US 5695937 A	09-12-1997
		US 5866330 A	02-02-1999
		AU 707846 B	22-07-1999
		AU 6561496 A	20-03-1997
		AU 7018896 A	01-04-1997
		CA 2185379 A	13-03-1997
		GB 2305241 A,B	02-04-1997
		IE 80465 B	12-08-1998
		JP 10511002 T	27-10-1998
		WO 9710363 A	20-03-1997
WO 9729211 A	14-08-1997	AU 2264197 A	28-08-1997

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82